

Evaluating Rare Variants in Complex Disorders Using Next-Generation Sequencing

Matthew Ezewudo · Michael E. Zwick

© Springer Science+Business Media New York 2013

Abstract Determining the genetic architecture of liability for complex neuropsychiatric disorders like autism spectrum disorders and schizophrenia poses a tremendous challenge for contemporary biomedical research. Here we discuss how genetic studies first tested, and rejected, the hypothesis that common variants with large effects account for the prevalence of these disorders. We then explore how the discovery of structural variation has contributed to our understanding of the etiology of these disorders. The rise of fast and inexpensive oligonucleotide sequencing and methods of targeted enrichment and their influence on the search for rare genetic variation contributing to complex neuropsychiatric disorders is the next focus of our article. Finally, we consider the technical challenges and future prospects for the use of next-generation sequencing to reveal the complex genetic architecture of complex neuropsychiatric disorders in both research and the clinical settings.

Keywords Human genetics · Genomics · Genetic architecture · Complex traits · Next-generation sequencing · Targeted enrichment · Single nucleotide variants · SNVs · Single nucleotide polymorphisms · SNPs · Structural variation · Copy number variants · CNVs · Complex neuropsychiatric disorders · Schizophrenia · Autism · Genetic disorders · Psychiatry

This article is part of the Topical Collection on *Genetic Disorders*

M. Ezewudo · M. E. Zwick (✉)
Department of Human Genetics, Emory University, Whitehead
Biomedical Research Building, Suite 301, Atlanta, GA 30322, USA
e-mail: mzwick@emory.edu

M. Ezewudo
e-mail: mezewud@emory.edu

Introduction

The individual and societal burdens of common, complex neuropsychiatric disorders are truly profound [1]. One of the major goals of contemporary biomedical research is to elucidate those disease mechanisms that underlie complex neuropsychiatric disorders like autism spectrum disorder (ASD) or schizophrenia (SZ). The hope is that an understanding of the pathogenesis of these disorders will enable the development of new treatments for those patients already affected, and new preventatives for those who are not. Because susceptibility to neuropsychiatric disorders is influenced by variation in both genes and environmental exposures, both genetic and epidemiological studies can help uncover novel disease mechanisms.

Our review focuses on what genetic studies of complex neuropsychiatric diseases have revealed about their genetic architecture, with a particular emphasis on studies of schizophrenia and autism. We divide the review into four main sections that reflect the technologies, experimental designs, and hypotheses tested in both recent and ongoing genetic studies of complex neuropsychiatric disorders. The first section discusses the recent history of human genetic studies, which have focused on the contribution of common variation to the risk of complex diseases. The second examines the contributions of exceptionally rare variants with large effects on disease risk. The third section addresses how the rapid development of next-generation sequencing and the targeted enrichment of eukaryotic genomes are contributing to studies of complex traits. Finally the last section focuses on challenges facing the application of next-generation sequencing in both research and clinical translational applications.

Common Genetic Variation and Complex Disease

For human genetic studies, the decade after the initial sequencing and analysis of a human reference genome has been a revolutionary one [2, 3]. The scaffold, a reference genome provided, allowed us to catalog one class of human genetic variation: single nucleotide polymorphisms (SNPs). Subsequent studies dramatically reduced the cost of genotyping genome-wide collections of hundreds of thousands of SNPs, while at the same time developing a map of the patterns of statistical correlation among common SNP variants, referred to as linkage disequilibrium, in the HapMap project [4, 5]. Furthermore, theoretical predictions suggested that a classic experimental design derived from epidemiology, a case-control association study, would have more statistical power than traditional genetic family-based linkage studies [6]. With these technologies in hand, genome-wide studies of complex disorders became feasible.

From this point, the conceptual framework and the types of experiments pursued were driven by both the availability of high-throughput genotyping platforms developed during the HapMap project and assumptions about the genomic architecture of common complex disorders [7]. The initial application of these technologies focused on an experimental design called the genome-wide association study (GWAS). The human genetic GWAS “industry” set out to test the hypothesis that common variants (those with >5 % frequency in the human population) with large effects contributed significantly to the risk of disease. The essential idea was that common disease-causing variants, which were expected to be found at an elevated frequency in cases as compared to matched controls, would either be genotyped directly or be in linkage disequilibrium with common SNPs, thereby allowing them to be discovered. This approach was successful in identifying many novel loci that contribute to a wide variety of complex diseases [8–10].

Unfortunately, the results of multiple genome-wide association studies of common neuropsychiatric disorders have been far more modest. Studies of schizophrenia (SZ) revealed only a few loci that exceed genome-wide levels of statistical significance, while the effect sizes of the variants uncovered were remarkably small [11–15]. Moreover, only a modest amount of the total heritability of SZ has been accounted for, in contrast to other complex traits, such as human height [16–18]. Similarly for autism, multiple GWAS identified a few loci of very small effect [19–21]. A subsequent meta-analysis suggested that finding any genetic variants with an odds ratio greater than 1.5 for autism is extraordinarily unlikely [22]. For neuropsychiatric disorders, therefore, the effect sizes of the variants identified have been disappointingly small, particularly when compared to GWAS of other complex human disease traits.

It is in fact the small effect size of common variants that is the most striking finding from nearly all genome-wide association studies of complex diseases. Together, these studies have soundly rejected the hypothesis that *common variants* with *large effects* underlie the vast majority of complex human diseases. Thus, while one can argue that the GWAS approach has been a success, these studies have revealed that the genetic architecture of most complex diseases is unlike that seen in cystic fibrosis or sickle cell anemia, where common alleles with very large effects account for most of the disease prevalence in human populations. This finding is particularly relevant for complex neuropsychiatric disorders like autism spectrum disorders and schizophrenia. Furthermore, although there are statistically compelling associations between common genetic SNPs and diseases [23, 24], single nucleotide polymorphisms (SNPs) alone are unable to account for all the genetic proportion of heritability in complex traits. The fact that a substantial proportion of the estimated heritability of these traits remains unexplained points to other classes of genetic variants that have yet to be discovered [25–27].

Rare Genetic Variation and Complex Disease

In retrospect, perhaps this outcome should have been less surprising. Theoretically, it has long been recognized that both common and rare variation likely contribute to the genetic architecture of complex traits [7, 28–35]. In addition, genome-wide association studies of common variants were pursued for the simple reason that technological advances made this experiment feasible. Direct sequencing of genomes to identify the contribution of rare variants in large numbers of patient samples faced daunting technological challenges and excessive costs that simply made such studies impractical.

While large-scale genotyping of SNPs for GWAS was underway, similar genome-wide technologies led to the discovery of widespread variations in copy number across the human genome [36–40]. This class of genetic variation, consisting of deletions and duplications larger than 100 kb, was surprisingly frequent. Although clinical geneticists had long recognized that cytologically visible, and usually much larger, chromosomal changes were associated with rare human diseases, the developing technologies allowed the discovery of smaller copy number variants (CNVs) that were not observable using classic cytological approaches. The role of this structural variation in human disease became an immediate focus [41, 42].

Soon after, the discovery of an elevated frequency of CNVs in patients with schizophrenia hinted at an explanation for the great heterogeneity of the disorder [15, 43–48]. Similar findings also came to light for autism [49–54], as well as for both schizophrenia and autism [55]. Nevertheless,

the apparently pathogenic CNVs discovered to date are in general large and very rare in the population, which means they alone are unable to explain all of the missing heritability.

Next-Generation Sequencing, Targeted Enrichment, and Complex Disease

Comprehensive sequencing of human genomes would no doubt be better at capturing the allelic architecture of complex diseases than the genotyping of common variants in GWAS or the detection of rare, large CNVs with methods like array comparative genomic hybridization, but even in the recent past this has been cost-prohibitive. Recent advances in next-generation sequencing (NGS), however, have increased throughput while decreasing costs, so this barrier is eroding quickly [56–63] see review in [64]. Combining NGS with methods that can enrich for portions of complex eukaryotic genomes has made it feasible to pursue other types of genetic variation underlying complex disease traits.

The initial application of these technologies focused on whole-exome sequencing, which involves sequencing the 1 % of the human genome that codes for proteins, in the context of diseases caused by mutations at single loci (so called Mendelian diseases) [65–67]. Application of these approaches to schizophrenia uncovered a role for *de novo* mutations in the etiology of the disorder [68, 69]. A more recent study suggested that many of the variants contributing to schizophrenia must be very rare and have yet to be discovered [70]. Whole-exome sequencing studies of autism also point to a role for *de novo* mutations in autism phenotypes [71–75]. Targeted studies of the X chromosome in males with autism, an attractive target given the 4:1 preponderance of males affected with the disorder, have revealed a number of putative autism susceptibility loci [76–79]. More recently, a combination of targeted enrichment of the X chromosome exome and next-generation sequencing identified the *AFF2* locus as having a significantly larger number of rare missense mutations in those with autism versus unaffected controls [80].

The clear message from all these studies is that exome sequencing can detect a broader allelic spectrum of complex neuropsychiatric disorders like schizophrenia and autism. As whole-genome sequencing becomes more and more cost effective, the field is bound to move toward this experimental design, which can reduce biases in ascertainment and make it possible to discover the full diversity of genetic variation.

Challenges Facing Next-Generation Sequencing and Complex Diseases

Next-generation sequencing and genomic enrichment technologies promise to detect both common and rare variants,

thereby giving us a better understanding of the genetic architecture of complex diseases, yet there are a number of substantial challenges facing the application of these technologies in both research, and ultimately, the clinic (see reviews in [81, 82]). We believe these challenges fall into three main categories: accurate identification of genetic variation, efficient analysis of next-generation sequencing data, and interpreting the functional effects of genetic variation.

Accurate Identification of Genomic Variation

Next-generation sequencing (NGS) technology platforms (Illumina, Roche 454, ABI SOLiD, Ion Torrent) have higher error rates in individual sequence reads than conventional Sanger sequencing. These errors could be systemic and significant enough to yield false-positive variant calls and associations, as well as obscure actual associations. Making even more errors possible are biases that arise in coverage: GC-rich genomic regions tend to have lower sequencing coverage. Further, enrichment technologies add another layer of possible errors, especially those methods that select sequences by hybridization to a complementary oligonucleotide. The existence of gene families and other repetitive regions imply that multiple genomic regions can be captured and enriched by a single oligonucleotide substrate targeted at a specific region. Finally, errors in mapping sequence against a human genome reference sequence can lead to the misidentification of genetic variants. This is of particular concern because the human genome reference sequence is an idealized genome, and undetected variation among individuals can lead to spurious outcomes in the mapping and identification of genomic variation [83].

The fact that the human genome is very large ($\sim 3 \times 10^9$ base pairs, or 3000 Mb) implies that extraordinary accuracy is necessary to identify variant sites. Even modest error rates of 1 %, for instance, could impugn the validity of association studies [84]. As a simple example, if we consider only SNPs, we expect approximately 3 million variant sites per genome. As shown in Table 1, in a scenario that surely underestimates the possible sources of error, unless algorithms calling variants sites are exceptionally accurate, the result will be an enormous number of false-positive findings.

Efficient Analysis of Next-Generation Sequencing Data

Assuming we are able to accurately identify variant sites, the next step is functionally annotating those same sites so we can focus on those most likely to contribute to disease. Genomic variation identified with NGS technologies needs to be annotated to establish its type, genomic region, the evolutionary conservation of its site, and whether it has prior characterization. A typical

Table 1 Expected number of errors in human whole-genome sequencing

Error rate	Expected number of errors (3000-Mb human genome)	Expected number of variant sites (per human genome)	Expected proportion of false-positive variant sites
1.0E-3	3,000,000	3,000,000	0.5
1.0E-4	300,000	3,000,000	0.1
1.0E-5	30,000	3,000,000	0.01
1.0E-6	3,000	3,000,000	0.001
1.0E-7	300	3,000,000	0.0001
1.0E-8	30	3,000,000	0.00001
1.0E-9	3	3,000,000	0.000001

whole-genome association study of a population would yield millions of variants, data that cannot realistically be characterized using public web genome browsers because of the huge effort involved. One early solution to this problem was the open source Sequence Annotator, or SeqAnt [85]. Researchers will have to rely increasingly on such high-performance annotation tools to analyze the sequencing data generated from large sequencing studies [86].

Interpreting the Functional Effects of Genetic Variation

Ultimately, the goal of association studies is to link genetic variants to phenotypes through statistical tests that show significant connections (effect sizes) between the discovered variant and the phenotype of the disorder. In those cases of Mendelian disorders, where single variants can account for a given disease phenotype, interpreting the functional effects of genetic variation is in many cases easier.

For complex neuropsychiatric disorders, on the other hand, the challenge is far more formidable [81]. When performing statistical testing of association of hundreds of thousands of variants in a genome-wide study, one immediately confronts the multiple testing problem, which is simply that, when performing a very large number of tests, the expected number of findings that exceed a nominal threshold of 0.05 will be substantial. Statistical methods like a Bonferroni correction or permutation can be used to control this issue, such that only very significant association signals are selected and false positives are reduced [87]; however, the extent to which false-negative findings are increased by these approaches remains unclear and difficult to determine.

Beyond this, the extent of genomic variation, perhaps much of it having no impact on the patient's phenotype, provides a stark challenge to interpreting the effects of

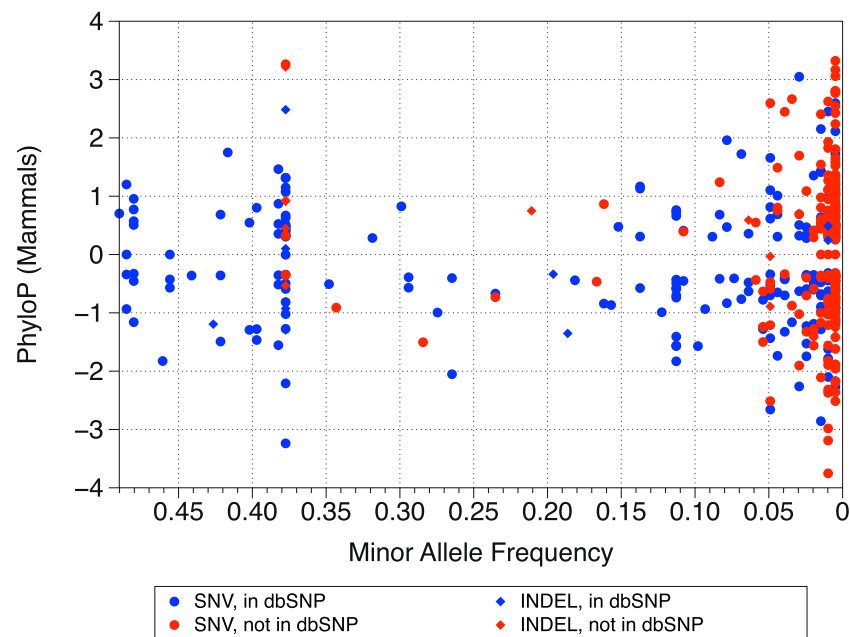


Fig. 1 Summary of single nucleotide variant (SNV) and insertion/deletion (indel) variation discovered at the FMR1 and AFF2 loci in males with autism spectrum disorder. The frequency of SNVs and indels (minor alleles) in cases is plotted against their level of evolutionary conservation. Population genetic theory predicts, and empirical data has now confirmed, that most genetic variation is rare. The observation of evolutionary conservation between species at a given site in the human genomes implies that genetic variation at this site is

deleterious to individual fitness and is therefore rapidly removed from the human population. Hence, it follows that disease-causing mutations are expected to be enriched among the class of rare variants found at highly evolutionary conserved sites in the human genome. Most common variation has already been discovered and exists in public databases like dbSNP (blue; circles and diamonds). In contrast, most of the rare variation at both loci was not contained in public databases (red; circles and diamonds)

genetic variation. Figure 1, for example, shows a summary of single nucleotide variants discovered through targeted sequencing of the genomic region containing the *FMR1* and *AFF2* loci in 144 boys with autism. A striking aspect of the figure is the enrichment for rare variants not seen before in public databases like dbSNP (Fig. 1). Population genetic models predicted that most variants will be rare, and recent genome-wide empirical studies have clearly established the vast excess of rare, previously unseen variants in human populations [88•, 89•]. As a result, to gain sufficient statistical power to identify genetic variants contributing to complex diseases, very large patient collections, on the order of 10,000 or more, are likely to be required.

Another approach that could help meet the statistical challenges of association studies is biological network and pathway analysis. Increasingly knowledge of biological pathways can enable statistical methods that leverage this information to discern associations between patterns of genetic variation and gene networks or pathways. Holistically testing for pathways and networks between genes across the groups being compared may improve power for associating genes to disease phenotype than the single variant comparison approach [90]. Still, these approaches presuppose knowledge of important pathways, and may not be the best way to uncover novel pathways or the action of mutant alleles that act outside of canonical pathways.

Finally, it is worth noting that the ultimate demonstration of causation will almost certainly fall beyond purely statistical methodologies. It may become necessary, and important if we are to understand fundamental disease mechanisms, to perform direct functional testing of variants *in vitro* or in model organisms *in vivo*. These experiments are far lower throughput than the original sequencing at the present time and likely represent a future bottleneck in our efforts to understand the genetic contribution to complex diseases. Furthermore, as we explore more deeply traits influenced by the actions of many genes, we will also need to more carefully examine the effects of environmental variation on the human traits of interest. In essence, DNA sequence and variation information is context-dependent, and to understand mechanisms of disease, we would ideally perform studies that can take into account both genomic variation information and putative environmental exposures.

Conclusions

The dramatic increase in the whole-exome and whole genome sequencing of large numbers of individuals has revealed more genetic variation between individuals than was previously suspected, as well as evidence for a higher incidence of rare and private variations in individuals within subpopulations. In a recent review on genetic variability

among humans, Olson emphasized that, although a number of different evolutionary and demographic forces act to influence human genomic variation, population genetics studies and, more recently, deep sequencing point to mutation-selection balance as having the greatest impact on the genetic predisposition to disease [91•].

GWAS studies of neuropsychiatric disorders have unequivocally shown that common variants with large effects do not underlie schizophrenia or autism. While statistical analyses of these complex disorders are consistent with the action of a very large number of common alleles of small effect, they are unable to account for the entire estimated heritability of the disorders. At the same time, while rare pathogenic CNVs can account for nearly all Mendelian forms of complex neuropsychiatric illnesses, because they are so rare in the general population, alone they cannot explain all of the missing heritability. Now, with the rapid advances and reduced costs of whole-genome sequencing, human geneticists will finally be able to more comprehensively uncover all classes of human genetic variation in large patient populations. Sifting through these enormous datasets will undoubtedly pose a stiff challenge for human geneticists, particularly given the tremendous heterogeneity and complexity underlying neuropsychiatric illnesses like autism and schizophrenia. There is little doubt that better integration of genomics with collaborative studies in physiology, biochemistry, and epidemiology is vital if we are to truly understand disease mechanisms and develop innovative methods of prevention and treatment for these devastating disorders.

Conflict of Interest M. Ezewudo: none; M.E. Zwick: grant from National Institutes of Health/National Heart, Lung, and Blood Institute, and consultant to Henry M. Jackson Foundation.

References

Papers of particular interest, published recently, have been highlighted as:

- Of importance
- Of major importance

1. Eaton WW, Martins SS, Nestadt G, et al. The burden of mental disorders. *Epidemiol Rev.* 2008;30:1–14.
2. Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. *Nature.* 2001;409:860–921.
3. Venter JC, Adams MD, Myers EW, et al. The sequence of the human genome. *Science.* 2001;291:1304–51.
4. Consortium IH. A haplotype map of the human genome. *Nature.* 2005;437:1299–320.
5. Consortium IH. A second generation human haplotype map of over 3.1 million SNPs. *Nature.* 2007;449:851–61.

6. Risch N, Merikangas K. The future of genetic studies of complex human diseases. *Science*. 1996;273:1516–7.
7. Zwick ME, Cutler DJ, Chakravarti A. Patterns of genetic variation in Mendelian and complex traits. *Annu Rev Genomics Hum Genet*. 2000;1:387–407.
8. Klein RJ, Zeiss C, Chew EY, et al. Complement factor H polymorphism in age-related macular degeneration. *Science*. 2005;308:385–9.
9. Consortium WTCC. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007;447:661–78.
10. Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. *Am J Hum Genet*. 2012;90:7–24.
11. O'donovan MC, Craddock N, Norton N, et al. Identification of loci associated with schizophrenia by genome-wide association and follow-up. *Nat Genet*. 2008;40:1053–5.
12. Stefansson H, Ophoff RA, Steinberg S, et al. Common variants conferring risk of schizophrenia. *Nature*. 2009;460:744–7.
13. Shi J, Levinson DF, Duan J, et al. Common variants on chromosome 6p22.1 are associated with schizophrenia. *Nature*. 2009;460:753–7.
14. Consortium IS, Purcell SM, Wray NR, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*. 2009;460:748–52.
15. Need AC, Ge D, Weale ME, et al. A genome-wide investigation of SNPs and CNVs in schizophrenia. *PLoS Genet*. 2009;5:e1000373.
16. Benyamin B, McEvoy BP, Gordon S, et al. Common SNPs explain a large proportion of the heritability for human height. *Nat Genet*. 2010;42:565–9.
17. Lee SH, Decandia TR, Ripke S, et al. Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. *Nat Genet*. 2012;44:247–50.
18. McQuillan R, Eklund N, Pirastu N, et al. Evidence of inbreeding depression on human height. *PLoS Genet*. 2012;8:e1002655.
19. Wang K, Zhang H, Ma D, et al. Common genetic variants on 5p14.1 associate with autism spectrum disorders. *Nature*. 2009;459:528–33.
20. Weiss LA, Arking DE, Daly MJ, Chakravarti A. A genome-wide linkage and association scan reveals novel loci for autism. *Nature*. 2009;461:802–8.
21. Anney R, Klei L, Pinto D, et al. A genome-wide scan for common alleles affecting risk for autism. *Hum Mol Genet*. 2010;19:4072–82.
22. Devlin B, Melhem N, Roeder K. Do common variants play a role in risk for autism? Evidence and theoretical musings. *Brain Res*. 2011;1380:78–84.
23. Manolio TA, Brooks LD, Collins FS. A HapMap harvest of insights into the genetics of common disease. *J Clin Invest*. 2008;118:1590–605.
24. Manolio TA. Genomewide association studies and assessment of the risk of disease. *N Engl J Med*. 2010;363:166–76.
25. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. *Nature*. 2009;461:747–53.
26. Maher B. Personal genomes: the case of the missing heritability. *Nature*. 2008;456:18–21.
27. Eichler EE, Flint J, Gibson G, et al. Missing heritability and strategies for finding the underlying causes of complex disease. *Nat Rev Genet*. 2010;11:446–50.
28. Falconer DS. The inheritance of liability to diseases with variable age of onset, with particular reference to diabetes mellitus. *Ann Hum Genet*. 1967;31:1–20.
29. Lande R. The maintenance of genetic variability by mutation in a polygenic character with linked loci. *Genet Res*. 1975;26:221–35.
30. Turelli M. Heritable genetic variation via mutation-selection balance: Lerch's zeta meets the abdominal bristle. *Theor Popul Biol*. 1984;25:138–93.
31. Barton NH, Turelli M. Evolutionary quantitative genetics: how little do we know? *Annu Rev Genet*. 1989;23:337–70.
32. Pritchard JK, Cox NJ. The allelic architecture of human disease genes: common disease-common variant or not? *Hum Mol Genet*. 2002;11:2417–23.
33. Barton NH, Keightley PD. Understanding quantitative genetic variation. *Nat Rev Genet*. 2002;3:11–21.
34. Schork NJ, Murray SS, Frazer KA, Topol EJ. Common vs. rare allele hypotheses for complex diseases. *Curr Opin Genet Dev*. 2009;19:212–9.
35. Visscher PM, Goddard ME, Derks EM, Wray NR. Evidence-based psychiatric genetics, AKA the false dichotomy between common and rare variant hypotheses. *Mol Psychiatry*. 2011;17:474–85.
36. Iafrate AJ, Feuk L, Rivera MN, et al. Detection of large-scale variation in the human genome. *Nat Genet*. 2004;36:949–51.
37. Sebat J, Lakshmi B, Troge J, et al. Large-scale copy number polymorphism in the human genome. *Science*. 2004;305:525–8.
38. Sharp AJ, Locke DP, McGrath SD, et al. Segmental duplications and copy-number variation in the human genome. *Am J Hum Genet*. 2005;77:78–88.
39. Redon R, Ishikawa S, Fitch KR, et al. Global variation in copy number in the human genome. *Nature*. 2006;444:444–54.
40. Perry GH, Tchinda J, McGrath SD, et al. Hotspots for copy number variation in chimpanzees and humans. *Proc Natl Acad Sci USA*. 2006;103:8006–11.
41. Stankiewicz P, Lupski JR. Structural variation in the human genome and its role in disease. *Annu Rev Med*. 2010;61:437–55.
42. Girirajan S, Campbell CD, Eichler EE. Human copy number variation and complex genetic disease. *Annu Rev Genet*. 2011;45:203–26.
43. Walsh T, McClellan JM, McCarthy SE, et al. Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science*. 2008;320:539–43.
44. Xu B, Roos JL, Levy S, et al. Strong association of de novo copy number mutations with sporadic schizophrenia. *Nat Genet*. 2008;40:880–5.
45. Stefansson H, Rujescu D, Cichon S, et al. Large recurrent microdeletions associated with schizophrenia. *Nature*. 2008;455:232–6.
46. McCarthy SE, Makarov V, Kirov G, et al. Microduplications of 16p11.2 are associated with schizophrenia. *Nat Genet*. 2009;41:1223–7.
47. Mulle JG, Dodd AF, McGrath JA, et al. Microdeletions of 3q29 confer high risk for schizophrenia. *Am J Hum Genet*. 2010;87:229–36.
48. Magri C, Sacchetti E, Traversa M, et al. New copy number variations in schizophrenia. *PLoS One*. 2010;5:e13422.
49. Sebat J, Lakshmi B, Malhotra D, et al. Strong association of de novo copy number mutations with autism. *Science*. 2007;316:445–9.
50. Kumar RA, KaraMohamed S, Sudi J, et al. Recurrent 16p11.2 microdeletions in autism. *Hum Mol Genet*. 2008;17:628–38.
51. Weiss LA, Shen Y, Korn JM, et al. Association between microdeletion and microduplication at 16p11.2 and autism. *N Engl J Med*. 2008;358:667–75.
52. Marshall CR, Noor A, Vincent JB, et al. Structural variation of chromosomes in autism spectrum disorder. *Am J Hum Genet*. 2008;82:477–88.
53. Glessner JT, Wang K, Cai G, et al. Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature*. 2009;459:569–73.
54. Levy D, Ronemus M, Yamrom B, et al. Rare de novo and transmitted copy-number variation in autistic spectrum disorders. *Neuron*. 2011;70:886–97.
55. Moreno-De-Luca D, Mulle JG, Kaminsky EB, et al. Deletion 17q12 is a recurrent copy number variant that confers high risk of autism and schizophrenia. *Am J Hum Genet*. 2010;87:618–30.
56. Bentley DR, Balasubramanian S, Swerdlow HP, et al. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*. 2008;456:53–9.

57. Wheeler DA, Srinivasan M, Egholm M, et al. The complete genome of an individual by massively parallel DNA sequencing. *Nature*. 2008;452:872–6.
58. Kidd JM, Cooper GM, Donahue WF, et al. Mapping and sequencing of structural variation from eight human genomes. *Nature*. 2008;453:56–64.
59. Lupski JR, Reid JG, Gonzaga-Jauregui C, et al. Whole-genome sequencing in a patient with Charcot-Marie-Tooth neuropathy. *New Engl J Med*. 2010;362:1181–91.
60. Fujimoto A, Nakagawa H, Hosono N, et al. Whole-genome sequencing and comprehensive variant analysis of a Japanese individual using massively parallel sequencing. *Nat Genet*. 2010;42:931–6.
61. Drmanac R, Sparks AB, Callow MJ, et al. Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science*. 2010;327:78–81.
62. Tong P, Prendergast JG, Lohan AJ, et al. Sequencing and analysis of an Irish human genome. *Genome Biol*. 2010;11:R91.
63. Consortium GP. A map of human genome variation from population-scale sequencing. *Nature*. 2010;467:1061–73.
64. Metzker ML. Sequencing technologies - the next generation. *Nat Rev Genet*. 2010;11:31–46.
65. Ng SB, Turner EH, Robertson PD, et al. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature*. 2009;461:272–6.
66. Choi M, Scholl UI, Ji W, et al. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc Natl Acad Sci USA*. 2009;106:19096–101.
67. Ng SB, Buckingham KJ, Lee C, et al. Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet*. 2010;42:30–5.
68. Xu B, Roos JL, Dexheimer P, et al. Exome sequencing supports a de novo mutational paradigm for schizophrenia. *Nat Genet*. 2011;43:864–8.
69. Girard SL, Gauthier J, Noreau A, et al. Increased exonic de novo mutation rate in individuals with schizophrenia. *Nat Genet*. 2011;43:860–3.
70. Need AC, McEvoy JP, Gennarelli M, et al. Exome sequencing followed by large-scale genotyping suggests a limited role for moderately rare risk factors of strong effect in schizophrenia. *Am J Hum Genet*. 2012;91:303–12.
71. O’Roak BJ, Deriziotis P, Lee C, et al. Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. *Nat Genet*. 2011;43:585–9.
72. Sanders SJ, Murtha MT, Gupta AR, et al. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature*. 2012;485:237–41.
73. O’Roak BJ, Vives L, Girirajan S, et al. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature*. 2012;485:246–50.
74. Neale BM, Kou Y, Liu L, et al. Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature*. 2012;485:242–5.
75. Iossifov I, Ronemus M, Levy D, et al. De novo gene disruptions in children on the autistic spectrum. *Neuron*. 2012;74:285–99.
76. Noor A, Whibley A, Marshall CR, et al. Disruption at the PTCHD1 Locus on Xp22.11 in Autism spectrum disorder and intellectual disability. *Sci Transl Med*. 2010;2:49ra68.
77. Filges I, Röthlisberger B, Blattner A, et al. Deletion in Xp22.11: PTCHD1 is a candidate gene for X-linked intellectual disability with or without autism. *Clin Genet*. 2011;79:79–85.
78. Chung R-H, Ma D, Wang K, et al. An X chromosome-wide association study in autism families identifies TBL1X as a novel autism spectrum disorder candidate gene in males. *Mol Autism*. 2011;2:18.
79. Piton A, Gauthier J, Hamdan FF, et al. Systematic resequencing of X-chromosome synaptic genes in autism spectrum disorder and schizophrenia. *Mol Psychiatry*. 2011;16:867–80.
80. Mondal K, Ramachandran D, Patel VC, et al. Excess variants in AFF2 detected by massively parallel sequencing of males with autism spectrum disorder. *Hum Mol Genet*. 2012. doi:10.1093/hmg/dds267.
81. Kiezun A, Garimella K, Do R, et al. Exome sequencing and the genetic basis of complex traits. *Nat Genet*. 2012;44:623–30.
82. Lyon GJ, Wang K. Identifying disease mutations in genomic medicine settings: current challenges and how to accelerate progress. *Genome Med*. 2012;4:58.
83. Rosenfeld JA, Mason CE, Smith TM. Limitations of the human reference genome for personalized genomics. *PLoS One*. 2012;7:e40294.
84. Glenn TC. Field guide to next-generation DNA sequencers. *Mol Ecol Resour*. 2011;11:759–69.
85. Shetty AC, Athri P, Mondal K, et al. SeqAnt: a web service to rapidly identify and annotate DNA sequence variations. *BMC Bioinforma*. 2010;11:471.
86. Ward LD, Kellis M. Interpreting noncoding genetic variation in complex traits and human disease. *Nat Biotechnol*. 2012;30:1095–106.
87. Aickin M, Gensler H. Adjusting for multiple testing when reporting research results: the Bonferroni vs Holm methods. *Am J Public Health*. 1996;86:726–8.
88. •• Keinan A, Clark AG. Recent explosive human population growth has resulted in an excess of rare genetic variants. *Science*. 2012;336:740–3. *This article characterizes how the explosive growth of human populations has impacted the patterns of human genetic variation, which need to be considered in studying the genetics of complex diseases and traits.*
89. • Tennessen JA, Bigham AW, O’Connor TD, et al. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science*. 2012;337:64–9. *This study shows an excess of rare functional variants in human populations and suggests that large sample sized will be required to associate rare variants with complex traits.*
90. Sun YV. Integration of biological networks and pathways with genetic association studies. *Hum Genet*. 2012;131(10):1677–86.
91. • Olson MV. Human genetic individuality. *Ann Rev Genomics Hum Genet*. 2012. doi:10.1146/annurev-genom-090711-163825. *A review that argues that rare genetic variation, whose dynamics are governed by mutation-selection balance, account for the majority of human genetic individuality.*